MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

22 **G** - 15

AN INTRODUCTION TO ESTIMATION THEORY FOR DYNAMICAL SYSTEMS

Fred C. Schweppe

AFESD-TDR-63-

28 August 1963

FCS:sm

LEXINGTON   MASSACHUSETTS

ABSTRACT

Estimation (filtering) theory for dynamical systems is reviewed. Emphasis is on time varying systems and nonstationary stochastic processes. The basic ideas for linear systems are presented in an intuitive manner using time domain techniques and the state variable concept. Both continuous and discrete time systems are discussed. A control problem and the principle of least squares curve fitting are related to the basic estimation problem. In addition to the presentation of fundamental principles for linear systems, brief discussions on a wide variety of related subjects are included in an appendix.
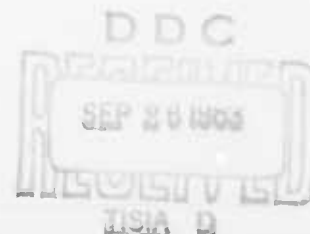
TABLE OF CONTENTS

1. INTRODUCTION

The following material is an introduction to the theory of estimation (or filtering) for nonstationary stochastic processes and time varying systems. The presentation has two main goals:

    1.) Provide physical insight and understanding of the basic principles.

    2.) Furnish guide posts for those who wish to delve deeper into the subject.

This dual intent divides the material into two parts. The main text develops the basic concepts along intuitive lines. The appendix supplements the basic ideas with cursory discussions on a wide range of related topics.

Section 2 defines the class of problems to be discussed and provides several examples. Section 3 considers the discrete time case in considerable detail. The discrete time (sampled data) case is emphasized as the basic ideas are easier to visualize and the general theory becomes most valuable in large scale systems which use sampled data and contain digital computers. Section 4 relates the theory of Section 3 to an equivalent problem in optimum control and the principle of least squares curve fitting. Section 5 gives the continuous time formulas along with some motivation.

2. PROBLEM STATEMENT

The development of the theory draws heavily on time domain techniques and the description of dynamical systems in terms of state variables. This general approach (like others such as frequency domain techniques and transfer functions) has both disadvantages and advantages but is especially appropriate for the present problem. In a heuristic sense, the state of a dynamical system is the minimum set of quantities which summarize enough information on the past of the system's input to determine the system's present and future output, assuming that the future inputs to the system are known. These quantities are called the state variables of the system. As an example, consider a passive electrical network containing resistors, capacitors and inductors. If the initial values of the currents in the inductors and the voltages across the capacitors are known, the output of the network at any specified future time can be calculated provided the inputs to the system are given. Thus, the state of the network at time t_0

1

is determined by the inductor currents and the capacitor voltages at time t_0.

Rather than present explicit discussions on the state variable concept, its use and implications, the development merely begins with systems represented in the state variable notation. The examples at the end of this section illustrate how some physical problems can be represented in this form.

In the main text, interest is confined to linear dynamical systems which can be described by systems of linear vector equations such as

$$\frac{d}{dt} X(t) = A(t)X(t) + B(t)U(t) \qquad (2.1)$$

$$Y(t) = H(t)X(t) \qquad (2.2)$$

or in discrete time

$$X(n\Delta) = \Phi(n\Delta)X((n-1)\Delta) + B(n\Delta)U(n\Delta) \qquad (2.3)$$

$$Y(n\Delta) = H(n\Delta)X(n\Delta) \qquad (2.4)$$

In these equations, X is the state vector,* U is the system input and Y is the system output. U and Y may be either vectors or scalars. The matrices $A(t), B(t), H(t)$ or $\Phi(n\Delta), B(n\Delta), H(n\Delta)$ determine the characteristics of the dynamical system. The variable, t, denotes time for continuous time systems, Eq. (2.1) and Eq. (2.2). For discrete time systems, Eq. (2.3) and Eq. (2.4), the variable, $n, n = 1, 2, \ldots,$ provides the time dependence while Δ is the time interval between "steps" of the discrete system. To simplify the discussion, Δ is assumed unity except in Section 5. Equations (2.1) thru (2.4) encompass models for a very wide range of linear, time varying dynamic systems. The major class of linear systems excluded (when X has a finite dimension) are those with pure time delay and distributed parameters.

It should be noted that the state space representation (in either continuous or discrete time) is not unique. Any nonsingular linear transformation of the state vector, $X(n)$, results in a new state vector with a corresponding $\Phi(n), B(n)$ and $H(n)$. These are merely different choices of "coordinate systems" in which the system is represented.

_____

* Vector and matrix quantities are denoted by capital letters. The corresponding lower case letter is used for the components of the matrix or in the scalar case. All vectors are column vectors. Transposition is denoted by a prime.

The desired theory is developed using the interplay of stationary, stochastic process and time varying linear systems. In particular, interest is centered on the vector, zero mean, "white noise" stochastic process. In the discrete time case, $W(n)$, is such a process if

$$E(W(n)W'(k)) = I \delta_{nk}$$

$$E(W(n)) = 0$$

where "E" denotes the expected value, δ_{nk} is the Kronecker delta and I is the unit matrix. For the continuous time case,

$$E(W(t)W'(t+\tau)) = I \delta(\tau)$$

$$E(W(t)) = 0$$

where $\delta(\tau)$ is the Dirac delta. In the discussions of the main text, no other assumptions on the actual probability distributions of the processes are made. Appendix A. 2 discusses the effect of assuming Gaussian distribution for the white noise processes.*

In all the following discussions, the same symbol is used for a random variable and a sample of a random variable. This abuse of notation should not lead to confusion.

There are two ways in which these white stochastic processes enter the problem; the system input may be random and/or measurement noise may be associated with the measurement of the system output. The exact model to be considered for the discrete time case is as follows. Define

$$X(n) = \Phi(n)X(n-1) + B(n)U(n) \tag{2.5}$$

where

_____

* In the literature, the term white noise often implies Gaussian distributions.

3

$X(n)$:   p-vector (the state)

$\Phi(n)$:   p x p state transition matrix

$B(n)$:   p x q matrix

$U(n)$:   zero mean, q-dimensional random process with

$$E(U(n)U'(m)) = I\delta_{nm}$$

$$R(n) = \text{p x p covariance matrix of } B(n)U(n)$$

$$= E\{B(n)U(n)U'(n)B'(n)\}$$

$$= B(n)B'(n).$$

Define

$$Z(n) = Y(n) + C(n)V(n)$$

$$= H(n)X(n) + C(n)V(n) \tag{2.6}$$

where

$Z(n)$:   r-vector

$H(n)$:   r x p matrix

$C(n)$:   r x r matrix

$V(n)$:   zero mean, r-dimensional random process with

$$E(V(n)V'(m)) = I\delta_{nm}$$

$$Q(n) = \text{r x r covariance matrix of } C(n)V(n)$$

$$= C(n)C'(n).$$

The random processes, $V(n)$ and $U(n)$, are uncorrelated; that is,

$$E(U(n)V'(m)) = 0 \quad \text{all n and m.}$$

In this model, $U(n)$ is the system excitation while $V(n)$ represents errors associated with the measurement of the system output. Thus interest is confined to stochastic processes which are either white or can be obtained by exciting a linear, time varying system with white noise. This includes all stationary stochastic processes with a rational power spectrum. From a mathematical point of view, Eq. (2.5) describes a pth order Markov process. This approach is an extension of

4

the pre-whitening filter concept used by Bode and Shannon in their derivation of the Weiner-Hopf frequency domain theory for stationary processes, see Ref. 1.

The model statement of Eqs.(2.5) and (2.6) is not complete as $X(1)$, the initial state of the system has not been specified. There are two approaches of interest.

1). Assume an a priori distribution for $X(1)$. That is; consider $X(1)$ a zero mean, random vector with covariance matrix, Q_0, where this random vector is uncorrelated with the U's and V's.

2). Consider $X(1)$ as an unknown parameter vector.

Fortunately, as far as the formulae to be developed are concerned, there is little difference between the two approaches and thus a choice between them need not be made here. In actual applications, the choice depends on the nature of the physical problem but is often of more philosophical than practical importance.

The model explicitly incorporates the following situations.

1). Deterministic system $(U(n) \equiv 0)$ observed in the presence of white noise.

2). Precisely measured stochastic process $(V(n) \equiv 0)$.

3). Stochastic process (or white noise driven deterministic system) observed in the presence of white noise.

The model, however, is actually applicable to a much wider range of physical problems. The case where the measurement noise is not white, but can be expressed in the form of Eqs.(2.1), (2.2) or (2.3), (2.4), is handled by appropriate re-definitions of the state variables, (see Example 2.3). Correlation between U and V is handled in a similar manner. Extension to nonzero, but known mean values for U and V can be treated by subtraction of the mean. Unknown, but parameterizable mean values can be incorporated into the dynamic system (see Example 2.4).

Define $\hat{X}(n)$ as the optimum estimate of $X(n)$ using $Z(k)$, $k = 1, \ldots, n$. The discussions of the main text are confined to linear estimation procedures; that is, when $\hat{X}(n)$ is a linear operation on the $Z(k)$, $k=1, \ldots, n$. In addition, the criteria of optimum is that of minimum variance. Let $\Sigma(n)$ be defined as

$$\Sigma(n) = E\{ (\hat{X}(n) - X(n)) (\hat{X}(n) - X(n))' \}$$

5

The optimum estimate is such that $\Sigma(n)$ is "smaller" than that obtainable from any other linear estimate. Although $\Sigma(n)$ is a matrix, it is such that it is possible to find a "smallest" $\Sigma(n)$. The optimum estimate, $\hat{X}(n)$, is unbiased; that is,

$$E(\hat{X}(n)) = X(n).$$

Appendix A. 2 discusses more general formulations of the problem which lead to the same formulas.

The problem to be considered in the main text is the determination of the linear operation on the $Z(k)$, $k=1, \ldots, n$ that results in the optimum estimate $\hat{X}(n)$, of $X(n)$.

Further define $\hat{X}(n/n-1)$ as the optimum estimate of $X(n)$ using $Z(k), k=1, \ldots, n-1,$ and

$$\Sigma(n/n-1) = E\left\{ (\hat{X}(n/n-1) - X(n))(\hat{X}(n/n-1) - X(n))' \right\}$$

An exact statement of the model for the continuous time case is as follows. The discrete time comments are also applicable here. Define

$$\frac{dX(t)}{dt} = A(t)X(t) + B(t)U(t)$$

where

$X(t)$:   p-vector (the state)

$A(t)$:   p x p matrix

$B(t)$:   p x q matrix

$U(t)$:   zero mean, q-dimensional random process with

$E(U(t)U'(t+\tau) = I\delta(\tau)$

$R(t) = B(t)B'(t)$

Define

$$Z(t) = Y(t) + C(t)V(t)$$
$$= H(t)X(t) + C(t)V(t)$$

6

where

$Z(t)$:   r-vector (the observations)

$H(t)$:   r x p matrix

$C(t)$:   r x r matrix

$V(t)$:   zero mean, r-dimensional random process with

$E(V(t)V'(t+\tau)) = I\delta(\tau)$

$Q(t) = C(t)C'(t)$

$V(t)$ and $U(t)$ are uncorrelated.

    The following examples illustrate how various problems can be put into the form of the general models. These examples are, of course, very simplified. A complete analysis of a realistic problem containing a periodically sampled continuous system is found in Ref. 2. Still other examples are found in Ref. 3.

Example 2.1

    Assume a radar is tracking a missile and obtains measurements of range at different times and that the measurement errors are samples from zero mean, (uncorrelated) random variables with variance $q(n)$. That is:

$$z(n) = r(n) + \Delta r(n)$$

where all quantities are scalars, $r(n)$ is the true range at time n, and $\Delta r(n)$ is discrete white noise. Assume that the true range behaves as a second degree polynomial in time. That is,

$$r(n) = r_1 + \dot{r}_1 n + \ddot{r}_1 \frac{n^2}{2}$$

where r_1, $\dot{r}_1$, and $\ddot{r}_1$, represent the range, range velocity and range acceleration at time n=1. To put this into the format of the general problem let:

7

$$\begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1/2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(n-1) \\ x_2(n-1) \\ x_3(n-1) \end{bmatrix}$$

Here $x_1(n)$ corresponds to the range $r(n)$ and x_2 and x_3 to the first and second time derivatives of the range, respectively. Then

$$z(n) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \end{bmatrix} + c(n)v(n)$$

or, in terms of the general formulation

$$\Phi(n) = \begin{bmatrix} 1 & 1 & 1/2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$H(n) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

$$B^2(n) = R(n) = 0$$

$$c^2(n) = Q(n).$$

For this example $X(1)$ could be considered either unknown or a random variable depending on the problem.

8

<u>Example 2.2</u>

Consider the following RLC network
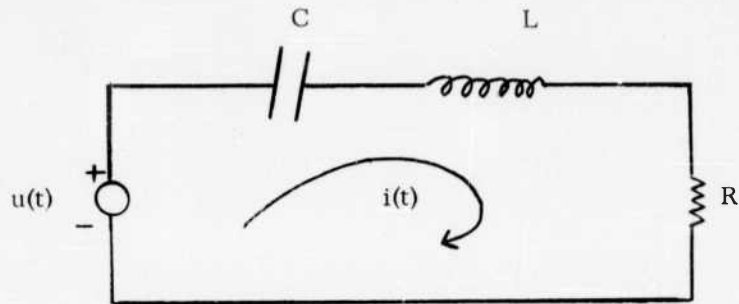
The switch is closed at t=0. The signal generator generates a voltage, u(t). Before the switch is closed, the current through L and the voltage across C are zero. Define

$$x_1(t) = \int_0^t i(t) \, dt = \text{charge on capacitor} .$$

Then

$$L \frac{d^2}{dt^2} x_1(t) + R \frac{d}{dt} x_1(t) + \frac{1}{C} x_1(t) = u(t).$$

Define

$$\frac{dx_1(t)}{dt} = x_2(t) = \text{current}.$$

Then

$$\frac{d}{dt} \begin{bmatrix} x_1(t) \\ \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \\ -\frac{1}{LC} & -R/L \end{bmatrix} \begin{bmatrix} x_1(t) \\ \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \\ 1/L \end{bmatrix} u(t)$$

9

If the voltage across R is desired,

$$z(t) = \begin{bmatrix} 0 & R \end{bmatrix} \begin{bmatrix} x_1(t) \\ \\ x_2(t) \end{bmatrix}.$$

Assume $u(t)$ is white noise of unit variance. Then, in terms of the general formulation,

$$A(t) = \begin{bmatrix} 0 & 1 \\ \\ -1/LC & -R/L \end{bmatrix}$$

$$B(t) = \begin{bmatrix} 0 \\ \\ 1/L \end{bmatrix}$$

$$H(t) = \begin{bmatrix} 0 & R \end{bmatrix}$$

$$R(t) = \begin{bmatrix} 0 & 0 \\ \\ 0 & 1/L^2 \end{bmatrix}$$

$$Q(t) = 0$$

A more realistic problem is obtained if additive white noise is assumed to corrupt the measurement of the voltage across R. In this case

$$z(t) = \begin{bmatrix} 0 & R \end{bmatrix} \begin{bmatrix} x_1(t) \\ \\ x_2(t) \end{bmatrix} + c(t)v(t)$$

10

and

$$Q(t) = c^2(t).$$

In either of these cases the initial conditions on $X(t)$ are assumed to be random variables.

Example 2.3

Consider Example 2.1 but assume the measurement error, $\Delta r(n)$, is the sum of a discrete white noise process plus a first order Markov process, $y(n)$. That is

$$\Delta r(n) = y(n) + cv(n)$$

where

$$y(n) = ay(n-1) + b\,u(n)$$

where $u(n)$ is discrete white noise of unit variance. Then the dynamical model is

$$
\begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \\ x_4(n) \end{bmatrix}
=
\begin{bmatrix} 1 & 1 & 1/2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & a \end{bmatrix}
\begin{bmatrix} x_1(n-1) \\ x_2(n-1) \\ x_3(n-1) \\ x_4(n-1) \end{bmatrix}
+
\begin{bmatrix} 0 \\ 0 \\ 0 \\ b \end{bmatrix} u(n)
$$

where $x_4(n)$ is $y(n)$.

$$
z(n) = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \\ x_4(n) \end{bmatrix}
+ cv(n)
$$

$$
R(n) =
\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b^2 \end{bmatrix}
$$

$x_4(1)$ is considered a random variable.

11

Example 2.4

In practice, measurement errors often have an unknown, nonzero mean, which can be assumed constant during the interval of measurements. Thus, consider Example 2.1 with

$$\Delta r(n) = cv(n) + \alpha$$

where α is a constant, but unknown, bias error. Then, in terms of the general formulation,

$$\begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \\ x_4(n) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1/2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(n-1) \\ x_2(n-1) \\ x_3(n-1) \\ x_4(n-1) \end{bmatrix}$$

$$z(n) = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \\ x_4(n) \end{bmatrix} + cv(n)$$

where $x_4(1)$ is considered an unknown. Obviously a wide class of bias errors can be handled in similar fashion. The major restriction is the ability to represent the bias error as a function of a finite number of unknown parameters.

3.   THE DISCRETE TIME CASE

The discrete time case will now be discussed. Although it is possible to obtain the desired solution directly in one step; the material is presented in four steps to separate the mathematical manipulations from the basic simplicity of the problem. The presentation is not entirely rigorous but rigorous derivations of the same formulas can

12

be found in many places, (see Appendix A. 2).

3.1 Scalar Case

A simple scalar case is considered first. Let the model be

$$z(n) = x(n) + c(n) v(n)$$

$$x(n) = x(n-1) = x \quad \text{(a constant)}$$

(3.1)

where $z(n)$, $x(n)$, $c(n)$ and $v(n)$ are all scalars. With respect to the general formulation of Section 2

$$\Phi(n) = 1$$

$$B(n) = 0$$

$$H(n) = 1.$$

The discussion is begun with x considered an unknown constant. Thus, Eq.(3.1) corresponds to making repeated, independent, measurements on a constant, where the variance of the measurement errors $c^2(n) = q(n)$, is time varying.

For the moment, ignore mathematical criteria and consider a "logical"[*] method of processing the $z(k)$, $k = 1, \ldots, n$ to obtain an estimate of the constant x. Since a linear operation is desired,

$$\hat{x}(n) = \sum_{k=1}^{n} w(k) z(k)$$

(3.2)

The problem is to pick the weighting coefficients, $w(k)$. Since the variance of the measurement errors is variable, the $w(k)$ should be chosen to weight the measurements with large variances the least. This can be done by making $w(k)$ proportional to $1/q(k)$ and Eq. (3.2) becomes,

$$\hat{x}(n) = \zeta(n) \sum_{k=1}^{n} z(k)/q(k)$$

(3.3)

---

[*] A different "logical" approach to the problem is discussed in Section 4.

The normalizing coefficient, $\zeta(n)$ can be fixed by assuming it is desirable for either,

1) $E(\hat{x}(n)) = x$ (i.e. $\hat{x}(n)$ is unbiased)

or

2) $\hat{x}(n) = x$ if no errors are present.

Since Eq. (3.3) can be rewritten

$$\hat{x}(n) = x \ \zeta(n) \sum_{k=1}^{n} \frac{1}{q(k)} + \zeta(n) \sum_{k=1}^{n} \frac{c(k)\,v(k)}{q(k)}$$

either of the above constraints gives,

$$\zeta(n) = \frac{1}{\displaystyle\sum_{k=1}^{n} 1/q(k)} \qquad . \qquad\qquad (3.4)$$

The variance, $\Sigma(n)$, of the estimate becomes

$$\Sigma(n) = E\left\{\hat{x}(n) - x\right\}^2$$

$$= \zeta^2(n)\, E\left[\sum_{k=1}^{n} \frac{c(k)\,v(k)}{q(k)}\right]^2$$

or from the independence of the $v(k)$,

$$\Sigma(n) = \zeta^2(n) \sum_{k=1}^{n} \frac{1}{q(k)}$$

or   (3.5)

$$\Sigma(n) = \zeta(n).$$

14

Thus, our logical data processing scheme is:

$$\hat{x}(n) = \Sigma(n) \sum_{k=1}^{n} \frac{z(k)}{q(k)} \qquad (3.6)$$

where

$$\Sigma(n) = \frac{1}{\sum_{k=1}^{n} \frac{1}{q(k)}} \qquad (3.7)$$

Now, consider the following mathematical problem. Find the $w(k)$ of Eq.(3.2) which minimize

$$L = E(x - \hat{x}(n))^2 = E\left\{x - \sum_{k=1}^{n} w(k) z(k)\right\}^2 \qquad (3.8)$$

subject to the constraint that

$$E(\hat{x}(n)) = x \quad . \qquad (3.9)$$

Taking expected values transforms the above into the problem of minimizing

$$\sum_{k=1}^{n} w^2(k) q(k)$$

subject to constraint

$$\sum_{k=1}^{n} w(k) = 1.$$

The solution, a straightforward exercise using an undetermined Lagrange multiplier, λ, is outlined below.

$$\frac{\partial}{\partial w_j} \left\{ \sum_{k=1}^{n} w^2(k)\, q(k) + \lambda \sum_{k=1}^{n} w(k) - \lambda \right\}$$

$$= 2w(j)\, q(j) + \lambda = 0 \qquad j = 1, \ldots, n.$$

Thus,

$$w_j = -\frac{\lambda}{2q(j)} \ .$$

From the constraint equation

$$-\frac{2}{\lambda} = \sum_{k=1}^{n} \frac{1}{q(k)} \ .$$

Thus, Eq. (3.6) and Eq. (3.7) actually gives a minimum variance, linear estimate subject to the constraint that the estimate is unbiased.

The above considered x an unknown constant. The other point of view wherein x is a zero mean random variable with variance q_0 will now be discussed. If a criteria analogous to Eq. (3.8) is used, it is desired to minimize

$$L = E\left(x - \sum_{k=1}^{n} w(k)\, z(k) \right)^2$$

$$= \left(1 - \sum_{k=1}^{n} w(k) \right)^2 q_0 + \sum_{k=1}^{n} w^2(k)\, q(k) \ . \tag{3.10}$$

Direct minimization of Eq. (3.10) with respect to the $w(k)$ gives the system of equations,

$$\frac{\partial L}{\alpha w_j} = -2\left(1 - \sum_{k=1}^{n} w(k) \right) q_0 + 2w(j)\, q(j) = 0 \qquad j = 1, \ldots, n,$$

16

whose solution is

$$w(j) = \frac{\Sigma(n)}{q(j)}$$

where

$$\Sigma^{-1}(n) = \frac{1}{q_0} + \sum_{k=1}^{n} \frac{1}{q(k)} \qquad (3.11)$$

and where $\Sigma(n)$ is the minimum value of L, of Eq. (3.10). Thus, for the case of x
a random variable, the results are essentially unchanged. The only difference is in
Eq. (3.11) where the variance q_0 appears as though an extra measurement of variance ·
q_0 has been made. Since this property also applies in the general case, it is obvious
why no great distinction is made between considering the initial conditions (x(1)) an
unknown parameter or a random variable with mean zero and variance q_0. A solution
based on x(1) an unknown, can be extended to x(1) a random variable by assuming
the existence of an extra measurement of variance q_0. The value of this hypothesized
measurement is the mean value of the random variable. Similarly, a solution based
on x(1), a zero mean random variable of variance q_0, becomes the solution for x(1)
unknown as $q_0 \rightarrow \infty$; that is, as the a priori distribution of x(1) "spreads out."

The estimation scheme can be written in a recursive formulation. From the
definitions of Section 2,

$$\hat{x}(n/n-1) = \text{estimate of } x(n) \text{ using data up to time } n-1,$$

$$= \hat{x}(n-1)$$

and

$$\Sigma(n/n-1) = E \left(\hat{x}(n/n-1) - x(n) \right)^2$$

$$= \Sigma(n-1) .$$

Then by simple algebraic manipulation, Eq. (3.6) and Eq. (3.7) give

17

$$\hat{x}(n) = \Sigma(n) \left\{ \Sigma^{-1}(n/n-1) \, \hat{x}(n/n-1) \right.$$
$$\left. + q^{-1}(n) \, z(n) \right\} \tag{3.12}$$

and

$$\Sigma(n) = \left\{ \Sigma^{-1}(n/n-1) + q^{-1}(n) \right\}^{-1} \tag{3.13}$$

Equation (3.12) states that $\hat{x}(n)$ is a weighted average of $\hat{x}(n/n-1)$ and $z(n)$ where the weights are proportional to the inverse of the variance.

The scalar case has been discussed in detail because the data processing scheme for the general vector case has the same form as those developed here and can be given the same motivation. That is, the estimate is a weighted average using weights proportional to the inverse of the variance. In the general case, the scalars become vectors and matrices and the formulas become more complex.

3.2 Linear Regression

The linear regression problem of classical statistics has the model

$$Z(n) = H(n) X(n) + C(n) V(n)$$

$$X(n) = X(n-1) = X .$$

Here $Z(n)$ and $V(n)$ are, r-vectors, X is a p-vector, $H(n)$ is a r x p matrix and $C(n)$ is a r x r matrix. This model corresponds to the general formulation of Section 2 with

$$\Phi(n) = I$$

$$B(n) = 0 .$$

This is more than a trivial exchange of matrices and vectors for scalars since measurements are made on some function, $H(n) X$, of X rather than on X itself.

It is shown in Appendix A. 5 that, when X is considered unknown, the optimum unbiased, linear estimate is given by

$$\hat{X}(n) = \Sigma(n) \sum_{k=1}^{n} H'(k)Q^{-1}(k) Z(k) \qquad (3.14)$$

where

$$\Sigma(n) = \left[\sum_{k=1}^{n} H'(k) Q^{-1}(k) H(k) \right]^{-1} \qquad (3.15)$$

$\Sigma(n)$ and $Q^{-1}(k)$ are assumed to exist, (see Appendix A. 6). The major difference between Eq. (3.14), (3.15) and Eq. (3.6) and (3.7) is the effect of the matrix $H(n)$. Thus, $1/q(k)$ from Eq. (3.7) is replaced by $H'(k) Q^{-1}(k) H(k)$. This effect can be motivated by assuming $H(n)$ is a square matrix with an inverse. For such a case, one can consider $\tilde{Z}(n)$ as the measurement quantity where

$$\tilde{Z}(n) = H^{-1}(n) Z(n)$$
$$= X + H^{-1}(n) C(n) V(n)$$

which is a vector version of the case in Section 3.1 with measurement error variance of $\tilde{Q}(n)$ where

$$\tilde{Q}^{-1}(n) = H'(n) Q^{-1}(n) H(n).$$

The proof in Appendix A. 5 simply shows that this same form also applies in the case where $H(n)$ is not invertible.

It can be shown that considering X a zero mean random variable with covariance matrix, Q_0, gives

$$\Sigma(n) = \left[Q_0^{-1} + \sum_{k=1}^{n} H'(k)Q^{-1}(k) H(k) \right]^{-1}$$

while Eq. (3.14) is unchanged.

19

As in Section 3.1, simple algebra manipulation gives

$$\hat{X}(n) = \Sigma(n)\left\{ \Sigma^{-1}(n/n-1)\hat{X}(n-1) + H'(n) Q^{-1}(n) Z(n)\right\} \qquad (3.16)$$

where

$$\Sigma(n) = \left\{ \Sigma^{-1}(n/n-1) + H'(n) Q^{-1}(n) H(n)\right\}^{-1} \qquad (3.17)$$

$$\Sigma(n/n-1) = \Sigma(n-1) \qquad (3.18)$$

$$\hat{X}(n/n-1) = \hat{X}(n \qquad (3.19)$$

The above recursive scheme is useful if a new estimate of X is desired each time a new measurement is made. However, in many applications only a final estimate at n = N, using all the data is desired. For such a problem it is better to rewrite the data processing scheme as

$$\Sigma^{-1}(n) = \Sigma^{-1}(n-1) + H'(n)Q^{-1}(n)H(n) \qquad (3.20)$$

$$\mathcal{H}(n) = \mathcal{H}(n-1) + H'(n) Q^{-1}(n) Z(n) \qquad (3.21)$$

$$\hat{X}(N) = \Sigma(N)\mathcal{H}(N). \qquad (3.22)$$

3.3 Dynamical System, No Excitation

The next step is the incorporation of a time variation in $X(n)$ where this variation is controlled by a deterministic dynamical system. The model is thus

$$Z(n) = H(n) X(n) + C(n) V(n)$$

$$X(n) = \Phi(n) X(n-1).$$

With respect to the general formulation of Section 2,

$$B(n) = 0.$$

20

This model is essentially the same as that of Section 3.2. The only difference is

$$\hat{X}(n/n-1) = \Phi(n)\,\hat{X}(n-1) \tag{3.23}$$

$$\Sigma(n/n-1) = \Phi(n)\,\Sigma(n-1)\Phi'(n) \tag{3.24}$$

The motivation for Eq. (3.23) is straightforward. Since the dynamical system is not excited, a state $X(n-1)$ at time n-1, results in a state $\Phi(n)X(n-1)$ at time n. Similarly, an estimate, $\hat{X}(n-1)$ should result in a predication of the next state given by Eq. (3.23). The actual proof is simple and is omitted. Equation (3.24) follows directly from Eq. (3.23) since

$$E\{(\hat{X}(n/n-1) - X(n)\,(\hat{X}(n/n-1) - X(n)\,)'\}$$
$$= E\{[\,\Phi(n)\,(\hat{X}(n-1) - X(n-1))]\,[\Phi(n)\,(\hat{X}(n-1) - X(n-1))]'\}$$
$$= \Phi(n)\,\Sigma(n-1)\,\Phi'(n)$$

Thus, for a dynamical system with no excitation,

$$\hat{X}(n) = \Sigma(n)\left\{[\Phi(n)\,\Sigma(n-1)\,\Phi'(n)\,]^{-1}\,\Phi(n)\,\hat{X}(n-1) \right. $$
$$\left. + H'(n)Q^{-1}(n)Z(n)\right\} \tag{3.25}$$

where

$$\Sigma(n) = \left\{[\,\Phi(n)\,\Sigma(n-1)\,\Phi'(n)]^{-1} + H'(n)\,Q^{-1}(n)\,H(n)\right\}^{-1} \,. \tag{3.26}$$

As in Section 3.2, these formulas can be written so that only one matrix inversion is required if only $\hat{X}(N)$ is desired.

3.4 Excited Dynamical System

The complete model of Section 2 is obtained by adding the final ingredient, a stochastic input to the dynamical system. Thus, the model is

21

$$Z(n) = H(n) X(n) + C(n) V(n)$$

$$X(n) = \Phi(n) X(n-1) + B(n) U(n).$$

As in Section 3.3,

$$\hat{X}(n/n-1) = \Phi(n) \hat{X}(n-1) \quad .$$

In this case, the system is excited, but by a zero-mean random vector. Thus, the best predication of the state ignores the effect of $B(n)U(n)$. The value of $\Sigma(n/n-1)$, however, is effected by the random excitation and it's calculation proceeds as follows:

$$\hat{X}(n/n-1) - X(n) = \Phi(n) \hat{X}(n-1) - \Phi(n) X(n-1) - B(n) U(n)$$

$$= \Phi(n)(\hat{X}(n-1) - X(n-1)) - B(n) U(n) .$$

Since $u(n)$ is uncorrelated with $\hat{X}(n-1)$ and $X(n-1)$,

$$\Sigma(n/n-1) = \Phi(n) \Sigma(n-1)\Phi'(n) + R(n) \tag{3.27}$$

where $R(n) = B(n) B'(n)$. Thus, for the general model

$$\hat{X}(n) = \Sigma(n) \left\{ \left[\Phi(n) \Sigma(n-1) \Phi'(n) + R(n) \right]^{-1} \Phi(n) \hat{X}(n-1) \right. \\ \left. + H'(n) Q^{-1}(n) Z(n) \right\} \tag{3.28}$$

where

$$\Sigma(n) = \left\{ \left[\Phi(n) \Sigma(n-1) \Phi'(n) + R(n) \right]^{-1} + H'(n) Q^{-1}(n) H(n) \right\}^{-1} \tag{3.29}$$

Equations (3.28) and (3.29) are, of course, the same form as that first developed in Section 3.1. The equations, however, can be rewritten to give a slightly different picture of the optimum process. Algebraic manipulation gives

$$\hat{X}(n) = \Phi(n) \hat{X}(n-1) + \Sigma(n) H'(n) Q^{-1}(n) \left\{ Z(n) - H(n) \Phi(n) \hat{X}(n-1) \right\} \tag{3.30}$$

22

The second term constitutes a correction, using the measurement $Z(n)$, to the old estimate, $\hat{X}(n/n-1) = \Phi(n)\hat{X}(n-1)$. This correction is weighted by the "ratio" of the variance of the old estimate to the variance of the new estimate. It is also possible to rewrite Eq. (3.29) as

$$\Sigma(n) = \Sigma(n/n-1) - \Sigma(n/n-1)H'(n)[Q(n) +$$
$$H(n)\Sigma(n/n-1)H'(n)]^{-1}H(n)\Sigma(n/n-1) \tag{3.31}$$

(This conversion is not obvious. It is developed in Ref. 2 using the formulas for the inversion of a partitioned matrix, see Ref. 4.) The second term of Eq. (3.31) shows how the variance is decreased by the new measurement. Equations (3.30) and (3.31) are only two of many possible variations of the basic formulas. References 2 and 3 contain still other forms.

Unfortunately, it is not possible in the general case $U(n) \neq 0$ to rewrite the formulas to require only matrix inversion (of order p) when only a final estimate is desired.

4. A RELATED CONTROL PROBLEM AND LEAST SQUARES CURVE FITTING

The principle of least squares curve fitting chooses one function from a specified class of functions so that the differences between it and some observed function is minimum in a least squares sense. An important control problem is to determine the particular input (or control law) for a system such that the system's output follows some prescribed path as closely as possible in the least squares sense. The solutions to these problems are closely related to the estimation results of Section 3. This property is discussed in Ref. 5 for the stationary, time invariant problem using Weiner-Hopf, frequency domain theory. Certain results for the time varying cases will be stated here. The actual derivations can be found in Refs. 6 and 7.[*]

The development of the desired relationship is begun with the scalar case of Section 3.1,

$$z(n) = x(n) + c\,v(n)$$
$$x(n) = x(n-1) = x$$

[*] This relationship between estimation and control is often called the duality principle.

23

Assume the estimation procedure is to choose the value of x that is closest to the measurements, z(n). Since the measure of distance should incorporate weighting to account for the different measurement variances, it is logical to choose x to minimize

$$\sum_{n=1}^{N} \frac{(x - z(n))^2}{q(n)} \quad .$$

It should not be surprising that this problem leads to the results of Section 3.1.

Now, consider the problem of choosing the initial conditions, X(1), for the dynamical system

$$X(n) = \Phi(n) X(n-1)$$

so that

$$L = \sum_{n=1}^{N} (Z(n) - H(n) X(n))' \, Q^{-1}(n) (Z(n) - H(n) X(n)) \tag{4.1}$$

is minimized. That is, choose the output, H(n) X(n), of the dynamical system which is closest to the observed quantities in the sense that Eq. (4.1) is minimized. If Q(n) is the unit matrix, the problem is that of least squares curve fitting. For general Q(n), it is a weighted least squares curve fit. The resulting values of X(n) equal those of Section 3.3.

For the general case of Section 3.4, it is desired to find the output, H(n) X(n), from the dynamical system

$$X(n) = \Phi(n) X(n-1) + B(n) U(n) \tag{4.2}$$

which is closest to Z(n). However, it is necessary to constrain the choice of the U(n), otherwise there exists a solution of Eq. (4.2) for which H(n) X(n) ≡ Z(n), all n. The formulas of Section 3.4 result when this constraint on U(n) is imposed by changing the criteria to the free minimization of

24

$$L = \sum_{n=1}^{N} (Z(n) - H(n)X(n))' \, Q^{-1}(n) \, (Z(n) - H(n)X(n))$$
$$+ \sum_{n=1}^{N} U(n)' \, R^{-1}(n) \, U(n)$$

(4.3)

Thus, for the general estimation problem, it is desired to determine the initial conditions $X(1)$ and the sequence $U(n)$ so that Eq. (4.3) is minimum. The minimization of Eq. (4.3) implies the minimization of the sum of a weighted least squares fit of $H(n)X(n)$ to $Z(n)$ and $U(n)$ to zero.

The $U(n)$ obtained from Eq. (4.3) can be considered an estimate of the realization of the random process that excited the system. In a similar manner, the $Z(n) - H(n)X(n)$ obtained from Eq. (4.3) can be considered an estimation of the realization of the measurement errors.

The problem of minimizing Eq. (4.3) is also a problem in control. That is, find the control signal, $U(n)$, which drives the dynamical system of Eq. (4.2) so that Eq. (4.3) is a minimum. The only difference is that $X(1)$ is usually given for the control problem.

A development of the estimation problem as one in weighted least squares can be obtained directly, for Gaussian $U(n)$ and $V(n)$, by using the statistical technique of Maximum Likelihood. This technique chooses as estimates of the unknown parameters (functions), the values which maximize the probability that the measurements, $Z(n)$, would actually be observed, see Ref. 8. This probability is maximized when Eq. (4.3) is minimum.

5. THE CONTINUOUS TIME CASE

The formulas for the continuous time problem can be obtained from the discrete time solution by letting the time interval between the steps go to zero as the number of steps goes to infinity. Difficulties arise because the continuous time white noise process has mathematical properties whose rigorous treatment requires a fair amount of mathematical sophistication (for example, infinite power). The following

25

discussions attempt only a heuristic motivation for the formulas. Reference 3 discusses the problem in more detail, but also without rigor. Reference 9 provides a rigorous basis using processes with independent increments but does not discuss the general problem considered here.

If Δ is the time between "steps," the discrete time model is

$$X(n\Delta) = \Phi(n\Delta)X((n-1)\Delta) + B(n\Delta)U(n\Delta) \tag{5.1}$$

$$Z(n\Delta) = H(n\Delta)X(n\Delta) + C(n\Delta)V(n\Delta) \tag{5.2}$$

$$E(U(n\Delta)U'(m\Delta)) = I\delta_{nm} \tag{5.3}$$

$$R(n\Delta) = B(n\Delta)B'(n\Delta)$$

$$E(V(n\Delta)V'(m\Delta)) = I\delta_{nm} \tag{5.4}$$

$$Q(n\Delta) = C(n\Delta)C'(n\Delta)$$

The corresponding continuous time model is

$$\frac{d}{dt}X(t) = A(t)X(t) + B(t)U(t) \tag{5.5}$$

$$Z(t) = H(t)X(t) + C(t)V(t) \tag{5.6}$$

$$E(u(t)u'(t+\tau)) = I\delta(\tau) \tag{5.7}$$

$$R(t) = B(t)B'(t)$$

$$E(V(t)V'(t+\tau)) = I\delta(\tau) \tag{5.8}$$

$$Q(t) = C(t)C'(t) \quad .$$

These two systems, discrete and continuous, must behave in a similar manner. That is, the discrete system may be considered the result of sampling the continuous system once every Δ time units. Similarly, as $\Delta \to 0$, $n \to \infty$ such that $n\Delta \to t$, the discrete system should converge in some statistical sense to the continuous system.

As a first step in relating the discrete and continuous cases, let

26

$$\Phi(n\Delta) = I + \Delta A(n\Delta)$$

Then Eq. (5.1) can be rewritten

$$\frac{X(n\Delta) - X((n-1)\Delta)}{\Delta} = A(n\Delta) X((n-1)\Delta)$$

$$+ \frac{B(n\Delta)}{\Delta} U(n\Delta)$$

(5.9)

Now, compare Eq. (5.9) with (5.5) and Eq. (5.2) with (5.6). As $\Delta \to 0$, $n \to \infty$, $n\Delta \to t$, it is not unreasonable to expect

$$A(n\Delta) \to A(t)$$

$$H(n\Delta) \to H(t)$$

$$X((n-1)\Delta) \to X(t)$$

$$Z(n\Delta) \to Z(t)$$

$$\frac{X(n\Delta) - X((n-1)\Delta)}{\Delta} \to \frac{d}{dt} X(t)$$

Furthermore, define

$$B(n\Delta) \to B(t)$$

$$C(n\Delta) \to C(t)$$

and therefore,

$$R(n\Delta) \to R(t)$$

$$Q(n\Delta) \to Q(t)$$

The more difficult concept is the behavior of the random processes, $\frac{U(n\Delta)}{\Delta}$ and $V(n\Delta)$ so that in the limit they can be associated with $U(t)$ and $V(t)$ respectively. The

27

required behavior is,

$$E[U(n\Delta) \, U'(n\Delta)] = \Delta I \qquad (5.10)$$

and

$$E(V(n\Delta) V'(n\Delta)) \;=\; \frac{I}{\Delta} \;. \qquad (5.11)$$

Since

$$E\left[\frac{U(n\Delta) \, U'(n\Delta)}{\Delta^2}\right] = \frac{I}{\Delta}$$

the correlation functions of the random processes, $\dfrac{U(n\Delta)}{\Delta}$ and $V(n\Delta)$, become the Dirac delta functions of Eqs.(5.7) and (5.8) as $\Delta \to 0$. Thus, Eqs.(5.10) and (5.11) furnish results that at least are consistent with Eqs.(5.7) and (5.8).

To motivate Eqs.(5.10) and (5.11), two simple scalar situations will be discussed. They are very special cases of the general problem but they illustrate the principle underlying the Δ behavior of Eqs.(5.10) and (5.11). The basic idea is that the behavior of the processes, $x(\Delta n)$ and $z(n\Delta)$, should not change as the number of "steps" taken in a fixed time interval is increased. Consider the scalar system

$$x(n) = x(n-1) + b\,u(n)$$

Now suppose the system is changed so that the state is transferred from time $n-1$ to time n in two steps instead of one. Let $u_2(n)$ be the stochastic process associated with this new system. Then

$$x(n-1/2) = x(n-1) + b\,u_2(n-1/2)$$

$$x(n) = x(n-1/2) + b\,u_2(n)$$

$$= x(n-1) + b\,[u_2(n) + u_2(n-1/2)]$$

28

If two systems are to be statistically equivalent, the variance of $u(n)$ and $(u_2(n) + u_2(n-1/2))$ must be equal. This requires

$$\frac{E(u^2(n))}{2} = E(u_2^2(n)) .$$

Similarly, if k steps, each of length $\frac{1}{k} = \Delta$ are taken,

$$E(u^2(n)) = \Delta E(u_k^2(n)) .$$

If this idea is extended to the time varying, vector case, Eq. (5.10) results. To motivate Eq. (5.11), consider the scalar system,

$$z(n) = x(n) + c v(n)$$

$$x(n) = x(n-1) = x$$

Now, change the system so that two observations at n and n-1/2 are made with measurement noise, $v_2(n)$. That is,

$$z(n) = x + c v_2(n)$$

$$z(n-1/2) = x + c v_2(n-1/2).$$

In order to obtain the same accuracy in estimating x from both the one and two observation systems, it is necessary that

$$E(v(n))^2 = \frac{E(v_2^2(n))}{2}$$

which when generalized to the actual problem leads to Eq. (5.11).

Hopefully, the preceding provides at least a feeling for why the recipe for using the discrete time formulas to obtain those for continuous time is:

29

Replace n by $n\Delta$

Replace $Q(n)$ by $\dfrac{Q(n\Delta)}{\Delta}$

Replace $R(n)$ by $\Delta R(n\Delta)$

Replace $\Phi(n\Delta)$ by $I + \Delta A(n\Delta)$.

Then let $n \to \infty$, $\Delta \to 0$, $n\Delta \to t$.

Substitution into Eq. (3.30) and re-arranging gives

$$\hat{X}(n\Delta) - \hat{X}((n-1)\Delta) = \Delta A(n\Delta) \hat{X}((n-1)\Delta)$$

$$+ \ \Sigma(n\Delta) H(n\Delta) \Delta Q^{-1}(n\Delta)\{Z(n\Delta) - H(n\Delta)[I + \Delta A(n\Delta)]\hat{X}((n-1)\Delta)\}$$

Dividing by Δ and letting $\Delta \to 0$, $n \to \infty$, $n\Delta \to t$; gives

$$\frac{d\hat{X}}{dt}(t) = A(t)\hat{X}(t) + \Sigma(t)H(t)Q^{-1}(t)\{Z(t) - H(t)\hat{X}(t)\} \tag{5.12}$$

as the equation governing the optimum estimate $\hat{X}(t)$. Equation (5.12) can be given a physical interpretation similar to that of Eq. (3.30).

To obtain $\Sigma(t)$, consider Eq. (3.31). Thus

$$\Sigma(n\Delta) - \Sigma(n\Delta/(n-1)\Delta) = \Sigma(n\Delta/(n-1)\Delta)H'(n\Delta)$$

$$\left[\frac{Q(n\Delta)}{\Delta} + H(n\Delta)\Sigma(n\Delta/(n-1)\Delta)H'(n\Delta)\right]^{-1} \tag{5.13}$$

$$H(n\Delta)\Sigma(n\Delta/(n-1)\Delta)$$

where

$$\Sigma(n\Delta/(n-1)\Delta) = [I + \Delta A(n\Delta)]\Sigma((n-1)\Delta)$$

$$[I + \Delta A(n\Delta)]' + \Delta R(n\Delta) \quad . \tag{5.14}$$

30

Substitution of Eq. (5.14) into (5.13), division by Δ and letting $\Delta \to 0$, $n \to \infty$, $n\Delta \to t$, gives

$$\frac{d}{dt}\, \Sigma(t) = A(t)\, \Sigma(t) + \Sigma(t)\, A'(t)$$

$$+ \; R(t) - \Sigma(t)\, H'(t)\, Q^{-1}(t)\, H(t)\, \Sigma(t)$$

(5.15)

as the equations governing the behavior of $\Sigma(t)$, the covariance of the error in the estimate $\hat{X}(t)$. The first two terms, $A(t)\Sigma(t) + \Sigma(t)A'(t)$ show how the variance changes when there is no more driving noise, $(R(t) = 0)$ and no more information is obtained (measurements of infinite variance, $Q(t) = \infty$, are made). The last two terms show how the driving noise increases the variance while the measurements decrease it.

As in Section 4, the continuous time estimation equations can be related to a corresponding control problem and the principle of least squares curve fitting.

6. DISCUSSION

A logical motivation for the estimation procedures has been attempted but mathematical optimality was prerequisite. As is well-known, a system optimized with respect to a single mathematical criteria is not necessarily a good design. In addition, there are many examples wherein engineering judgment has produced a system whose performance, with respect to the criteria, is close to that of the mathematically optimized system. There are, however, two excellent reasons for considering optimum procedures.

> 1) They furnish an excellent reference point for system feasibility studies and to determine whether or not an existing system can be appreciably improved.

> 2) An already derived solution frees engineering judgment for more important problems, or at least provides an initial design which can be remodeled to meet specific needs.

The second point is becoming increasingly important as system complexity grows and engineering solutions become correspondingly less obvious.

31

The formulation of the theory presented here is based on time domain techniques, state variables and dynamical systems. However, there are other formulations, such as frequency domain techniques and classical statistics. With respect to frequency techniques, the time domain approach has the big advantage that nonstationary problems can be easily formulated and, most important, the results are obtained in a form suitable for direct computer mechanization. The time and frequency domain techniques provide different, but equally valuable insights into the behavior of the optimum process. For stationary cases, the best approach depends entirely on the problem, (see Appendix A. 15). The basic difference between the final formulations presented here and classical statistics is the employment of dynamical systems. In many engineering applications the dynamical system formulation is natural and is an extremely valuable addition. However, there exist many important problems in which such a formulation is an unnecessary complication and the classical approaches (partially illustrated by Section 3. 2) are far more appropriate.

Thus in summary, it must be emphasized that the optimum theory developed here is not the ideal approach to all estimation problems. However, it is a very powerful tool, especially when tempered with engineering judgment.

APPENDIX

The main text has presented some basic concepts underlying estimation techniques for time varying systems and processes. There is a great abundance of other material in this area, some only minor mathematical frills, but much of fundamental and practical importance. In order to expose at least some of the terminology, this appendix contains very succinct discussions on certain of these concepts. Some of the references are to engineering literature, but many are of necessity to mathematically orientated presentations. The references are not a complete bibliography in any sense, but they should provide a basis from which a good understanding can be built.

A. 1 Historical Background

The history of the basic principle can be traced back to the work on least squares of Gauss and Legendare. From this original work the theory has evolved

32

through many different paths to its present state. The statisticians, economists, surveyors, natural scientists, and astronomers, all developed various aspects of the theory, often independently of each other. The engineers entered the picture with Weiner's work, see Ref. 10, and proceeded with their own evolution, often again re-deriving earlier results. No attempt will be made to delineate this evolution. The present development used the more classical approaches to arrive at a final formulation which was first derived and investigated by Kalman, see Refs. 3, 6, and 7.

A.2 Other Formulations and Criteria

There are many different ways to derive and interpret the basic formulas. A few of these are briefly discussed.

Section 4 indicates that the problem can be considered one in control. Thus, all of the techniques useful in optimal control theory, such as Dynamic Programming, Calculus of Variations, and Pontryagin's Maximum Principle, can be employed to derive and extend the basic results. In addition, the various concepts from optimal control theory of the Hamiltonian, canonical forms and the Hamilton-Jacobi equations can also be employed, see Ref. 3.

A derivation of the formulas can be done in a nice mathematical manner using projection theory in Hilbert space. This is done in Ref. 6 to derive the formulas for the continuous time case directly.

A convenient technique simply assumes the form of the optimum solution and proves it's optimality using a generalized Schwartz inequality, see Ref. 11. The resulting formulas are more general than those of Section 3.4. The proof in Appendix A.14 uses a modification of this technique.

The proof in Appendix A.14 minimizes the main diagonal terms of the error covariance matrix. However, the proof actually states that the difference between any nonoptimum error covariance matrix and the optimum error covariance matrix is a positive definite matrix and this implies that the error ellipsoid (or ellipsoid of concentration) for the optimum data processing scheme (see Appendix A.8) is entirely contained within any other possible error ellipsoid, see Ref. 12. This is another way of stating that the optimum estimate of a linear operation on $X(n)$ is simply the same linear operation on $\hat{X}(n)$, see Ref. 2.

33

When X(1) is considered unknown, it is often possible to reduce the error variance by allowing a biased estimate, see Ref. 13. At the present time, however, this possibility appears to be of mostly theoretical interest.

For Gaussian random processes many additional interpretations can be given the optimum estimate. It provides the conditional expectation of $X(n)$ given all available data. (Reference 3 develops the entire theory from this point of view.) The formulas can thus be shown optimum with respect to many other than mean square criteria, see Refs. 14, 15. The estimate is a maximum likelihood estimate, (see Section 4). It is also the estimate that contains the most information about $X(n)$, (see Appendix A. 3).

The estimation formulas require knowledge of only the mean and variance (first and second moments) of the various stochastic processes. In practice, it is often true that no other information on the stochastic processes is available. In such instances, there is no way of distinguishing the actual processes from Gaussian processes. This principle is related to the concept of "strict sense" and "wide sense" as discussed in Ref. 9.

A. 3 Relation to Information Theory

Optimum estimation theory has a basic relationship to information theory. For example, the matrix Σ^{-1} is defined as the information matrix (or Fisher information matrix) and shown to be directly related to the definition of information, see Ref. 16. If the quantities $H(n) Q^{-1}(n) Z(n)$ are considered as information, Eqs. (3. 20) and (3. 21) show that both the information and the information matrices are additive for independent measurements (i. e., $U = 0$). This is a property to be expected from information theory.

A. 4 Extrapolation and Interpolation

Only estimates of the present state of the system have been considered. That is, given $Z(n)$, $n = 1, \ldots, N$, it is desired to find an estimate of $X(N)$. Extrapolation is the problem of estimating $X(k)$, $k > N$ while interpolation is the problem of estimating $X(k)$ $k < N$.

Extrapolation is a simple extension of the ideas presented, in fact it was done (without proof) in the derivations of Sections 3. 3 and 3. 4. Interpolation is not difficult

34

but somewhat tricky. In the discrete time case, it can be done using the basic ideas presented and extension to the continuous case by a limiting process is probably possible. However, interpolation often appears to be best handled by looking at the problem as one in control (see Section 4).

A.5 Derivation of Eq. (3.14) and Eq. (3.15)

Equations (3.14) and (3.15) are derived by assuming a solution and proving it to be optimum. Appendix A.2 discusses other approaches.

It is desired to find the optimum $W(k)$ in

$$\hat{X}(n) = \sum_{k=1}^{n} W(k) Z(k) \qquad (A-1)$$

where

$$Z(k) = H(k) X + C(k) V(k) \qquad (A-2)$$

$$X = X(k) = X(k-1)$$

$\hat{X}(n)$ is constrainted to be an unbiased estimate of X. Thus

$$\sum_{k=1}^{n} W(k) H(k) = 1 \qquad (A-3)$$

Define

$$\tilde{\Sigma} = \left(\sum_{k=1}^{n} H'(k) Q^{-1}(k) H(k) \right)^{-1} \qquad (A-4)$$

and

$$W_k = \tilde{\Sigma} \left[H'(k) Q^{-1}(k) + \Theta(k) \right] \qquad (A-5)$$

Using Eqs. (A-3) and (A-4)

$$\tilde{\Sigma} \sum_{k=1}^{n} \Theta(k) H(k) = 0 \qquad (A-6)$$

35

Now

$$\Sigma(n) = E\{(\hat{X}(n) - X)(\hat{X}(n) - X)'\}$$

$$= E\{(\sum_{k=1}^{n} W(k) C(k) V(k))(\sum_{k=1}^{n} W(k) C(k) V(k))'\}.$$

Using the independence of the $V(k)$

$$\Sigma(n) = \tilde{\Sigma}\{\sum_{k=1}^{n} (H'(k) Q^{-1}(k) + \Theta(k)) Q(k)(H'(k)Q^{-1}(k) + \Theta(k))'\} \tilde{\Sigma}$$

Using Eqs. (A-6), (A-4)

$$\Sigma(n) = \tilde{\Sigma} + \tilde{\Sigma}\{\sum_{k=1}^{n} \Theta(k) Q(k)\Theta'(k)\} \tilde{\Sigma} \qquad (A-7)$$

Now, $\tilde{\Sigma}$ and $Q(k)$ are positive definite matrices. Thus, the main diagonal terms of the second term of Eq. (A-7) are positive for all $\Theta(k)$ other than $\Theta(k) = 0$. Since it is desired to minimize $\Sigma(n)$, (see also Appendix A. 2), the best choice for $\Theta(k)$ is zero. Thus,

$$\tilde{\Sigma} = \Sigma(n)$$

$$W(k) = \Sigma(n) H'(k) Q^{-1}(k)$$

which are the desired equations.

A. 6 Existence of Matrix Inverses

The formulas in the main text employ many matrix inversions. In some cases the matrices may be singular and the inverses do not exist, (see also Appendix A. 7). There are two basic situations depending on whether it is a singular covariance matrix, or singular information matrix. (If the inverse of the information matrix exists, it is a covariance matrix, see Appendix A. 3).

A singular information matrix means not enough information is available to estimate the parameters of interest. For example, the expression

$$\sum_{k=1}^{n} H'(k) Q^{-1}(k) H(k)$$

from Section 3.2 is the information matrix resulting from making n, r-dimensional measurements Z(k) on a p-dimensional unknown vector. If p is greater than n times r, this information matrix is singular as there are more unknowns than measurements. A singular information matrix occurs during the first stages of an estimation problem when the initial conditions are considered unknown parameters. For the case $U(n) = 0$, the problem can be handled easily by using Eqs.(3.20) and (3.21) and simply waiting until enough information has been obtained to allow inversion. The general case of $U \neq 0$ can be handled by waiting until enough information has been obtained and then combining all of the measurements into one vector, Z, of suitably increased dimension.

The inability to invert a covariance matrix implies a singular distribution of the random variables. For example, consider the random vector X where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

If $x_1 = x_2$, the distribution is singular and the covariance matrix of the vector X cannot be inverted. The presence of such a singular distribution usually justifies a re-appraisal of the system model but in some instances, a singular distribution may have to be handled. The classical (statistics) approach reduces the dimension of the vector, see Ref. 17. Thus, in the above example, X is replaced by a 2-dimensional vector. A systematic approach to this problem using the concept of a generalized inverse is given in Ref. 3.

A.7 Computational Considerations

The theoretical formulation is only part of the problem. An equally important aspect is the mechanization of the formulas. One of the beauties of the theory is the way in which the formulas (discrete time) can be directly implemented on a digital computer. Complex logic is not needed and storage requirements are small if the recursive formulas are employed. Mechanization requires only addition and multiplication (and matrix inversion). However, the choice of the particular formulas to be programmed is not necessarily straightforward and can constitute an important part of the problem. As discussed in Section 3, there are many different forms for the optimum estimator.

As an extreme example of the effect of formulation on efficiency, consider the case when $U = 0$ and where only one estimate using N vector measurements is desired. The formulation of Eq. (3.16), Eq. (3.17) requires N matrix inversions while Eqs.(3.20), (3.21) and (3.22) require only one.

In many applications, estimates of only certain states of the system are desired as in Example 2.4, where the state, x_4, corresponding to the bias error may be of no interest. It is possible to write formulas which explicitly provide only the desired states. (Reference 18 has detailed discussions on this and other problems for the case of $U = 0$). These formulas are more complicated than those which estimate the full state but may result in computational savings in that lower-order matrix inversions are required. This ability to trade complex equations for a reduction in size of an inverted matrix is a common occurrence. For example, Eq. (3.31) requires the inversion of a $r \times r$ matrix (r = dimension of measurement) while Eq. (3.29) is much simpler, but requires inversion of a $p \times p$ matrix (p = dimension of state). The choice between the two formulations depends on the problem and the relative size of p and r. Reference 4 discusses many ways and formulations for matrix inversions.

Unfortunately, even if efficiency is not important, the problem may not be straightforward as numerical round-off errors can prove disastrous, especially in the matrix inversion. A matrix may not be singular in the sense of Appendix A.6, but if it is badly ill conditioned, numerical inversion techniques can introduce large errors. Of course, specialized matrix inversion techniques may be employed, see Ref. 4.

38

However, it is often better to rephrase the problem by changing the state variables (the coordinate system) or by modifying the basic system model. This question of numerical accuracy must be emphasized as it is not an uncommon problem.

Reference 3 considers continuous time problems.

A.8 Error Ellipsoid and Ellipsoid of Concentration

Consider a Gaussian, zero mean, r-dimensional vector Y, with covariance matrix Σ. In the r-dimensional space of this vector, a contour of constant probability is given by

$$Y' \Sigma^{-1} Y = \text{constant}$$

which is the equation for an ellipsoid in r-dimensions. When the random vector Y is associated with the errors in an estimate, the resulting ellipsoid is often called the error ellipsoid.

For the non-Gaussian case, the ellipsoid of concentration replaces the error ellipsoid. This ellipsoid defines a random vector with uniform distribution over the ellipsoid and covariance matrix equal to that of the original non-Gaussian distribution.

Reference 8 is a good reference.

A.9 Regular, Sufficient, Efficient, Consistent and Admissible Estimates

For the general nonlinear problem, there are various ways in which estimates can be classified and five such groupings are: regular, sufficient, efficient, consistent and admissible.

Regularity refers to the boundedness of certain functions and their partial derivatives, see Ref. 8.

Sufficiency refers to the ability of one estimate to convey all of the pertinent information contained in many measurements, see Ref. 8. An example of this is found in Section 3 where $\hat{X}(N-1)$ contains all of the pertinent information in $Z(n)$, $n = 1, \ldots, N-1$ with regard to estimating $\hat{X}(N) \cdot \hat{X}(N-1)$ is a sufficient estimate.

Efficiency is a measure of how well an estimate is performing. The standard of comparison for this measure is the fundamental Cramer-Rao or Information inequality

39

which provides a lower bound on the error variance, see Refs. 8 and 16. For the linear Gaussian case, the estimates developed in Sections 3 and 5 are efficient in that the error variance equals that provided by the Cramer-Rao inequality.

A consistent estimate converges in probability to the unknown parameters as the number of measurements goes to infinity, see Ref. 8.

Admissibility applies to problems for which no one estimate is optimum for all values of the unknown parameters, see Ref. 12. An admissible estimate cannot be improved on at all values of the unknown parameters. Admissibility is thus a type of optimality.

A. 10 Observability, Controllability, Stability

Observability and controllability are recently coined terms for the classification of different types of systems. Controllability refers to the ability to drive a system from one state to another, while observability refers to the ability to calculate all the states of the system from observations on just certain states, see Refs. 3 and 19. As a very heuristic example, a system is not observable if the information matrix is not invertible.

Stability is, of course, a well-known term. The optimum data processing schemes may or may not be stable and this may or may not be a desirable property. Stability can be related to observability and controllability.

A. 11 Nonlinear Problems

A fairly general class of nonlinear problems can be written in the following form:

$$X(n) = F\{X(n-1), n, U(n)\}$$

$$Z(n) = H(n)X(n) + C(n)V(n)$$

where $U(n)$ and $V(n)$ are discrete, white random processes. For this general class of problems there is little reason to expect solutions that are optimum in any general sense. However, there are a wide variety of techniques which can provide useful and practical answers.

40

The maximum likelihood technique can be used. This requires the solution of a system of nonlinear equations. For $U(n) = 0$ and Gaussian $V(n)$, these equations can often be solved by Newton's method wherein the linear theory developed in Section 3 is incorporated in an iteration loop, see Ref. 20. Maximum likelihood is the most popular for solving nonlinear problems, but there are a variety of other important techniques such as BAN estimates (Best Asymptotically Normal) see Ref. 21, recursive re-linearization techniques, Ref. 22, the methods of moments, Ref. 8, and minimum Chi-square, Ref. 8. For the general case, maximum likelihood techniques become very difficult even for Gaussian random variables. However, a very reasonable processing scheme can be based on the method of least squares; that is, by looking at the estimation problem as one in least squares as discussed in Section 4 and using the minimization of Eq. (4.3) as the criterion. Unfortunately, there is little work available on the behavior of such estimates.

The "matched filter" concept of communication and radar technology can be interpreted as solving the maximum likelihood equations for the case $U = 0$, see Ref. 23.

A.12 Uncertain Variances and Dynamics

All of the preceding discussions have made the critical assumption that the variances of the stochastic processes and the structure of the dynamical systems are known. Unfortunately, there is no dearth of problems where either or both of these assumptions are extremely naive. The performance of the estimator is usually not critical with regard to the assumed variances of the measurement errors as the data processing procedure is merely performing a weighted average. As long as the weights are self consistent (for example, the estimate is unbiased), small errors in the relative magnitude in the weights should not be important. The effects of errors in the dynamical structure can be far more critical. The stability of the estimator is important in many instances. For example, with an unstable estimator a dynamical error such as the neglect of a small bias measurement error can cause unbounded effects on the estimator's performance. Thus, it is sometimes expedient to employ a system model which results in a stable estimator. On the other hand, uncertainties in the dynamical structure of the measurement noise may be unimportant if the filter is unstable as the estimator may be asymptotically efficient (see Appendix A.13)

41

Reference 24 handles dynamical uncertainties by comparing the performance of a simple estimator which is insensitive to the uncertainties with the performance of the optimum under the assumption the dynamics are known exactly.

When uncertainties are present, it is often desirable to use the actual measurements to at least partially resolve them. If the dynamics are known but the variances are not, it is often possible to estimate the variances of the measurement errors from the data itself. If the dynamics are not known, it is possible to use hypothesis testing techniques to check the validity of an assumed dynamical model against some class of alternate dynamical models. Unfortunately, the application of these ideas to the general case often lends to both mathematical complexities and horrendous computational problems and few results are presently available. However, the statisticians have made extensive studies on the special case when $U = 0$. This theory is often titled "The General Linear Hypothesis," see Refs. 16 and 17. Reference 25 is an example of what can be done in a special case.

Uncertainties offer unequalled opportunities to exploit ingenious designs which can be called adaptive systems.

A. 13 Asymptotic Efficiency

The efficiency of an estimate is measured relative to the Cramer-Rao or Informative inequality (see Appendix A. 9). Certain estimates which are not efficient for finite sample size become efficient as the number of measurements become infinite. These are called asymptotically efficient estimates, see Ref. 8.

In nonlinear problems (see Appendix A. 11) which have no optimum estimator for finite sample size, asymptotically efficient estimates (such as maximum likelihood, BAN, and recursive re-linearization) sometimes exist and such asymptotic efficiency is often used as a criterion in the choice of an estimation technique. Asymptotic efficiency is also a useful concept in linear problems. For example, in certain cases it can be shown that an estimation procedure based on the assumption of white measurement noise is asymptotically efficient even if the measurement noise is not actually white, see Ref. 26.

42

A.14 Fixed Memory Estimators

All of the preceding estimation techniques have had growing memories; that is, the new measurements are combined with all of the past data. However, in many instances, the uncertaintities in the system (see Appendix A.12) make it desirable to ignore the early measurements. As an illustration consider Example 2.1. The assumption that the missile behavior is a second degree polynomial in time may be a valid approximation for only a few seconds while the total tracking time may be several minutes. Thus, given $Z(k)$, $k = 1, \ldots, n$, it is often desired to find, $\hat{X}(n, n-\tau)$, the optimum estimate using $Z(k)$, $k = n-\tau, \ldots, n$ where τ is the memory length of the estimation scheme.

Such fixed memory filters go thru an initial transient when $n < \tau$ during which the growing memory theory applies. The formulas for the steady state operations can, of course, be obtained by employing those for an estimator operating on just $Z(k)$, $k = n-\tau, \ldots, n$. However, a fixed memory estimator can be written in the following form which is often computationally far more superior,

$$\hat{X}(n, n-\tau) = F_1(n) \hat{X}(n-1, n-1-\tau)$$
$$+ F_2(n) Z(n) + F_3(n) Z(n-1-\tau) .$$

Reference 27 gives the explicit formulas for a special case.

Appendix A.15: Relation to Weiner-Hopf Theory

Weiner-Hopf theory is a technique for obtaining optimum estimators for stationary stochastic processes by the use of frequency domain techniques. See Refs. 1, 10, 28, 29, and 30. As stated in Section 6, the time domain is generally far superior to the frequency domain for nonstationary problems. However, for the stationary case, the choice depends primarily on the problem and somewhat on the designer's personal preference. This subject is briefly discussed.

Consider Eqs. (5.12) and (5.15). For stationary systems which have reached steady state conditions, the variance of the estimate is constant with time. Thus, Eq. (5.15) becomes the algebraic equation,

$$A\Sigma + \Sigma A' + R - \Sigma H' Q^{-1} H \Sigma = 0 \qquad \text{(A.15-1)}$$

and Eq. (5.12) becomes the system of time invariant differential equations

$$\frac{d\hat{X}(t)}{dt} = A\hat{X}(t) + \Sigma H Q^{-1} \{ Z(t) - H\hat{X}(t) \} \qquad \text{(A.15-2)}$$

The optimum estimator is thus obtained by solving Eq. (A.15-1) for Σ. This solution is equivalent to the spectral factorization required in the Weiner Hopf theory, (see Ref. 3, for more detail).

The Weiner-Hopf frequency domain theory is actually more general than the time domain theory presented here as it requires no restriction to Markov processes; that is, to stochastic processes obtained by exciting linear systems of the form of Eq. (2.1) or Eq. (2.3) with white noise. In addition, the time domain techniques depend on the choice of the state variable representation; that is, the coordinate system in which the system is defined. In this sense, the frequency domain techniques are coordinate free, a property which can prove quite valuable. On the other hand, the use of an explicit coordinate system often provides a better understanding of the process's behavior.

For continuous time systems, the nature of the problem determines whether the frequency or time domain representations provides the best basis for synthesis of the system. However, discrete time systems are often implemented on digital computers and for such cases, time domain solutions provide the answer directly in the required format.

To illustrate the relationship between the time and frequency domain approaches, consider the following scalar problem.

$$z(t) = x(t) + cv(t)$$

where

$$\dot{x}(t) = -x(t) + u(t).$$

44

In terms of frequency domain techniques, cv(t) has a power spectral density given by

$$S_v(\omega) = q = c^2$$

and x(t) has a power spectral density given by

$$S_x(\omega) = \frac{1}{1+\omega^2} \quad .$$

If $G(\omega)$ denotes the transfer function of the optimum filter, then

$$G(\omega) = \frac{-\sqrt{q} + \sqrt{1+q}}{\sqrt{1+q} + i\sqrt{q}\,\omega} \quad .$$

This result is derived in Refs. 10, 29, and 30. Now, consider the corresponding time domain approach. In terms of the general formulation,

$$A(t) = -1$$
$$B(t) = 1$$
$$H(t) = 1$$
$$C(t) = c$$
$$R(t) = 1$$
$$Q(t) = q$$

Let $\Sigma = \sigma$, a scalar. Then Eq. (A.15-1) becomes,

$$-2\sigma + 1 - \frac{\sigma^2}{q} = 0$$

or since σ must be positive,

$$\sigma = -q + \sqrt{q^2 + q}$$

45

and Eq. (A.15-2) becomes

$$\frac{d\hat{x}(t)}{dt} = (-1 - \frac{\sigma}{q}) \hat{x}(t) + \frac{\sigma}{q} z(t),$$

a system whose transfer function is

$$\frac{\sigma/q}{i\omega + \frac{\sigma}{q} + 1}$$

which after substitution and rearrangement becomes the $G(\omega)$ obtained by the Weiner - Hopf theory.

REFERENCES

1. H. W. Bode and C. E. Shannon, "A Simplified Derivation of Linear Least Squares Smoothing and Prediction Theory," Proc. IRE, 38 (1950), p. 417-425.

2. D. J. Sakrison, "Optimum Incorporation of External Observations with a Marine Inertial Navigation System," Technical Report No. 289, Lincoln Laboratory, M. I. T., Nov. 30, 1962.

3. R. E. Kalman, "New Methods and Results in Linear Prediction and Filtering Theory," Technical Report 61-1, Research Institute for Advanced Studies, Baltimore, Md., (Nov. 1960).

4. E. Bodewig, Matrix Calculus, (North-Holland Publishing Co., 1956).

5. G. C. Newton, Jr., L. A. Gould, J. F. Kaiser, Analytical Design of Linear Feedback Controls, (John Wiley and Sons, 1957).

6. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," J. Basic Engr. (ASME Trans.), 83 D (1961).

7. R. E. Kalman, " A New Approach to Linear Filtering and Prediction Problems," J. Basic Engr. (ASME Trans.), 82 D (1960) p. 35-45.

8. H. Cramer, Mathematical Methods of Statistics (Princeton University Press, 1946).

9. J. L. Doob, Stochastic Processes, (John Wiley and Sons, 1953).

10. N. Wiener, The Extrapolation, Interpolation and Smoothing of Stationary Time Series, (John Wiley and Sons, 1949).

11. U. Grenander and M. Rosenblatt, Statistical Analysis of Stationary Time Series, (John Wiley and Sons, 1957).

12. D. Fraser, Nonparametric Methods in Statistics, (John Wiley and Sons, 1957).

13. L. LeCam, "On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes' Estimates," University of California Publications in Statistics, 1, No. 11, pp 277-330. University of California Press, Berkeley, 1953.

14. S. Sherman, "Non-Mean-Square Error Criteria," IRE Trans. on Information Theory, IT-4, pp 125-126; September 1958.

15. T. R. Benedict and M. M. Sondhi, "On a Property of Wiener Filters," Proc. IRE, 45, pp 1021-1022, July 1957.

16. S. Kullback, Information Theory and Statistics, (John Wiley and Sons, 1959).

17. T. W. Anderson, An Introduction to Multivariate Statistical Analysis, (John Wiley and Sons, 1958).

18. R. E. Wengert, "Techniques of Error Analysis and Redundant Data Adjustment," Guidance System Analysis, Defense Systems Dept., Syracuse, N. Y. (April 9, 1962).

19. R. E. Kalman, Y. C. Ho, and K. S. Narendra, "Controllability of Linear Dynamical Systems," 1, No. 2, (February 1961). Contributions to Differential Equations.

20. I. I. Shapiro, "The Prediction of Ballistic Missile Trajectories from Radar Observations," Technical Report No. 129, Lincoln Laboratory, M. I. T., Feb. 27, 1957.

21. T. S. Ferguson, "A Method of Generating Best Asymptotically Normal Estimates," Ann. Math. Stat., 29, pp 1046-1062, (1959).

22. L. A. Gardner, Jr., "A Recursive Estimation Procedure for Certain Nonlinear Regression Problems - Scalar Single Parameter Case," Lincoln Laboratory, M. I. T. Report, MS-888, May 6, 1963.

23. E. J. Kelly, I. S. Reed, and W. L. Root, "The Detection of Radar Echoes in Noise," J. Soc. Indust. Appl. Math, 8, No. 2 (June 1960).

24 D. J. Sakrison, "Comparison of Linear and Nonlinear Estimation Methods for Signals Containing Unknown Parameters," Lincoln Laboratory, M. I. T. Report to ap...

25. T. W. Anderson, "The Choice of the Degree of a Polynomial Regression as a Multiple Decision Problem," Mathematical Statistics, Columbia University, March 2, 1961.

26. M. Rosenblatt, "On the Estimation of Regression Coefficients of a Vector-Valued Time Series with a Stationary Residual," A. Math. Stat., 3d Berkeley Symposium.

27. M. Blum, "Fixed Memory Least Squares Filters Using Recursive Methods," IRE Trans. on Information Theory, IT-3 , pp 178-182, Sept, 1957.

28. J. L. Laning and R. H. Battin, Random Processes in Automatic Control (McGraw-Hill, 1956).

29. W. B. Davenport, Jr., W. L. Root, An Introduction to the Theory of Random Signals and Noise, (McGraw-Hill, 1958).

30. Y. W. Lee, Statistical Theory of Communication, (John Wiley and Sons, 1960).

DISTRIBUTION LIST

Division 2

C. R. Wieser
S. H. Dodd
F. C. Frick

Group 21

J. H. Halberstein
C. W. Uskavitch

Group 22

V. A. Nedzel
W. I. Wells

A. W. Armenti
R. J. Bergemann
L. F. Cianciolo
D. F. Clapp
E. L. Eaton
J. J. FitzGerald
A. Freed
L. A. Globus
A. A. Grometstein
J. H. Hartogensis
R. C. Holland
C. Hopkins
M. M. Klein
E. Korngold
W. Z. Lemnios
L. M. Meixsell
L. J. Murray
B. Nanni
C. E. Nielsen
J. H. Pannell
E. W. Pike
D. Reiner
J. Rheinstein
J. Salerno
B. J. Schafer
A. F. Smith
J. S. Strano
H. Sussman
J. L. Vernon

Group 28

J. A. Arnow
J. F. Nolan

C. R. Arnold
M. Athans
T. C. Bartee
M. Canon
P. L. Falb
L. A. Gardner, Jr.
L. Jones, III
L. Kleinrock
H. K. Knudsen
H. J. Kushner
R. Lacoss
J. B. Lewis
J. Moses
D. Schneider
F. C. Schweppe(20)
J. M. Winett
P. E. Wood

Group 41

Harrison Jones

Group 62

K. Jordan
M. Levin

Group 65

H. L. Vantrees

Hq. ESD (Maj. S. Jeffery ESRC)
L. G. Hanscom Field
Bedford, Mass.

Hq. ESD (Lt. J. Egan)
L. G. Hanscom Field
Bedford, Mass.